SCALITY
ZENKO
Freedom & control across Hybrid IT and Multi-Cloud
A Scality White Paper
October 2019
I. Introduction: Hybrid IT and Multi-Cloud Storage ........................................... 3
II. Previous Challenges for Multi-Cloud Architectures ........................................ 4
III. Key Requirements: Hybrid IT & Multi-Cloud Storage .................................... 5
IV. Scality Zenko – Multi-Cloud Data Controller .................................................. 6
V. Extensible Storage Locations and Multi-Cloud Location Control ....................... 7
   Public Clouds ....................................................................................................... 7
   Private Clouds and On-Premises Storage .......................................................... 7
VI. Unified Interface ............................................................................................... 9
VII. Federated Security Management ..................................................................... 10
VIII. Data Storage in Native Cloud Format ............................................................ 11
IX. In-Band & Out-of-Band Data and Metadata Updates ........................................ 12
X. Metadata-Driven, Policy-Based Data Management ......................................... 13
XI. Metadata Search ............................................................................................. 14
XII. Cloud-based Management Portal : Zenko Orbit ............................................ 16
Summary .............................................................................................................. 17
I. Introduction: Hybrid IT and Multi-Cloud Storage

For many enterprises, data is created and stored for the most part in on-premises storage solutions. However, use cases for Hybrid IT storage and even multi-cloud storage are emerging in many industries. Hybrid IT is the combination of on-premises resources with a public cloud service. Multi-cloud is the abstraction and use of multiple public cloud services with or without on-premises resources which themselves might be architected as a private cloud.

There are many examples of Hybrid IT and multi-cloud in today’s IT environment. For example, in the Media and Entertainment industry, the creation of content in on-premises private clouds leveraging object storage has become prevalent, but the use of public cloud services for content distribution (CDN), transcoding, and indexing is now also a requirement. In the Financial Services industry, on-premises storage still provides the best security, control and performance, but public clouds such as Amazon Glacier, Azure Archive Storage, and Google Coldline Storage offer compelling solutions for off-site compliance and long term archival. For a broad set of other industries, value-added cloud services for data management and analytics such as AWS Elastic Map Reduce, Azure Video Indexer, and Google Cloud Dataflow provide functionality that is essentially unattainable for the average enterprise to implement in their own data centers.
II. Previous Challenges for Multi-Cloud Architectures

Building applications and architectures that are truly multi-cloud has been difficult in the past given the differences between APIs and other management interfaces across public clouds and on-premises resources, not to mention the gateway products that sometimes connect the two. For example, Amazon S3 obviously supports the Amazon S3 API, which has become the de facto object storage API, but Azure Blob Storage and Google Cloud Storage do not. Dealing with multiple APIs and management interfaces has always been anathema for application developers who want to avoid the vendor lock-in of a single API as well as the added development, support, and time costs associated with supporting multiple APIs.

Additionally, while there are products that enable the movement and tiering of data from on-premises storage to public clouds, many of these offerings block the full advantages of cloud services by maintaining a proprietary data format that cannot be accessed outside of the specific product. Maintaining the native format of the storage location — in other words, storing data in Amazon S3 in Amazon format; storing data in Azure Blob Storage in Azure format; etc. — makes the data available for native cloud services that enable organizations to leverage the full range of services from each cloud for analytics, business intelligence, compute bursting, indexing, and more. Being able to use any service from any cloud at any time enables enterprises to maximize the value of their data through the full power and value-add of any cloud service.

What's needed is a solution that fundamentally changes data storage and data management. A solution that provides enterprises with the full power and flexibility of on-premises storage as well as all the storage and data management capabilities of any public cloud. Only with such a solution will enterprises gain the freedom and control to maximize the value of their data.
III. Key Requirements: Hybrid IT & Multi-Cloud Storage

Whether implementing a Hybrid IT or a multi-cloud storage architecture, a number of key capabilities are required:

- Extensible storage location capability that supports not only leading public cloud storage services (Amazon S3, Microsoft Azure Blob Storage, Google Cloud Storage, and others), but also leading on-premises, scalable object storage systems (such as Scality RING)

- Unified front-end interface for applications that virtualizes and simplifies the complexities and differences between the various storage locations – public cloud or on-premises

- Federated Security Management model that eliminates the complexity of cloud security across multiple clouds

- Data storage in native cloud format (non-obfuscated, non-single-vendor-proprietary) for open access and use by any application or value-added cloud service

- Support for both in-band as well as out-of-band data and metadata updates thus ensuring applications can access and modify data regardless of location while maintaining eventual consistency

- Scalable, easy-to-use meta data search that aggregates both system-and user-created metadata across storage locations

- Intuitive creation and management of business policies to control data workflows (replication, lifecycle, expiration, etc.) in and between storage locations
IV. Scality Zenko – Multi-Cloud Data Controller

A multi-cloud data controller is the key to unlocking the value of data and the power of cloud storage through total data transparency. A policy-based, metadata-driven, multi-cloud data controller enables a broad range of use cases including Cloud Media Workflows, Cloud Disaster Recovery for Data, Best of Breed Cloud Service, IoT Data Federation, and more.

With the first introduction of Scality Zenko into the open source community in the summer of 2017, Scality delivered a unique platform positioned to support multi-cloud use cases. In 2018, both the open source and enterprise versions of Zenko include advanced features including:

- Extensible storage locations and multi-cloud location control
- Unified interface
- Federated security management
- Data storage in native cloud format
- In-band and out-of-band data and metadata updates
- Meta data-driven, policy-based data management
- Metadata search
- Cloud-based management portal
V. Extensible Storage Locations and Multi-Cloud Location Control

Zenko supports extensible storage locations, as well as the concept of multi-cloud location control. The extensible storage locations allow new or additional storage locations (public clouds, on-premises solutions, etc.) to be added to a Zenko deployment over time for added flexibility and reduced vendor lock-in. As of GA in September 2018, Zenko supports the following storage locations:

Public Clouds
- Amazon S3 (and by extension lifecycle transition to Amazon Glazier)
- Azure Blob Storage (all storage tiers)
- Google Cloud Storage (all storage tiers)
- Digital Ocean
- Wasabi
- Others to be added in the future

Private Clouds and On-Premises Storage
- Scality RING
- Others to be added in the future
Multi-cloud location control enables storage in multiple clouds simultaneously as well as more advanced policies for data management across these multiple clouds.

In the Amazon S3 API, a Bucket contains a “location-constraint” sub-resource, as metadata. This is used to associate a Bucket with its S3 region (as of this writing, there are 14 regions, for example us-east-1, eu-west-1 others around the world). When a location is associated with a Bucket, all object PUT requests to that Bucket will thereafter be stored in that S3 region (with the Bucket itself managed in the Zenko namespace). The objects themselves are stored in the underlying cloud’s native format to preserve open access.

Zenko extends the Amazon S3 concept of location-constraint to provide control to non-Amazon S3-compatible public clouds as well on-premises private storage clouds. For example, Zenko applies the location-constraint concept to Azure Blob Storage such that data can be stored in specific Azure regions as well as in specific storage tiers (hot, cool, archive). multi-cloud location control applies to any of the storage locations Zenko supports – public cloud or on-premises.
VI. Unified Interface

Zenko leverages the Amazon S3 API for its north-bound application interface. By adopting a single, de facto standard API for object storage, Zenko provides application developers with a single, unified endpoint through which data can be stored, retrieved, and searched across all locations.

Zenko gives application developers a unified way to write applications without having to manage multiple APIs or protocols.
VII. Federated Security Management

For multi-cloud architectures, federation of the underlying security and identity models is a key requirement because each public cloud offers its own variation of identity management for users, groups, RBAC, etc. Zenko leverages the Amazon Identity and Access Management (IAM) model to unify security management across all clouds.

Zenko extends the powerful and well-known IAM model to non-AWS public clouds such as Azure and Google, as well as on-premises storage solutions such as Scality RING. Additionally, Zenko’s implementation of IAM can be federated with traditional enterprise security services such as Microsoft Active Directory (AD) and LDAP using the well-known SAML 2.0 identity provider solutions to map AD concept to IAM. This powerful capability ensures authentication/Single Sign-On (SSO) integrated with proven security services.
VIII. Data Storage in Native Cloud Format

As discussed previously, enterprises are actively using many value-added cloud services such as AWS Elastic Map Reduce, Azure Video Indexer, and Google Cloud Dataflow to analyze, index, search, or somehow process their data. Enterprises want the freedom and control to copy or move data between on-premises and public cloud or between multiple public clouds in order to access the appropriate cloud service for any given project. Zenko maintains the native data format for any given storage location, thus preventing lock-in to a vendor-proprietary format, thus ensuring that enterprises are not blocked from using other services to process data.

Many of today’s gateway products store data in opaque, proprietary formats rendering that data unusable by cloud services.
IX. In-Band & Out-of-Band Data and Metadata Updates

In a multi-cloud architecture, one or more applications may need to access data in any given storage location, and that access may not be directly through – or in-band – a single interface. For example, when storing data in an on-premises location as well as in Amazon S3 and Azure Blob Storage, what if an application accesses and modifies data directly in Amazon S3 – hence Out-of-Band? How does the multi-cloud data controller understand these changes and achieve eventual consistency?

Zenko supports both in-band and out-of-band updates. When used in-band, applications write to Zenko and Zenko updates the appropriate storage locations. When used Out-of-Band, an application can write directly to a storage location that Zenko is managing and the storage location can send metadata updates to Zenko such that Zenko is aware of data changes even though those changes were not made through Zenko.
X. Metadata-Driven, Policy-Based Data Management

Key to Zenko’s value proposition is managing metadata across all of its managed storage locations. This metadata can be used by Zenko’s policy-based data management engine to manage data workflows such as 1:1 replication, 1:Many replication, expiration, lifecycle, tiering and more.

Zenko’s data management policies are expressed using the Amazon S3 de facto standard XML-based configuration rules. Multiple policies can be expressed for each configuration with rules attached to Buckets. Rules may also be filtered on specific objects in the Bucket based on a name-prefix or a set of object “tag” values. For example:

- **Expiration**: expire objects automatically – objects starting with the prefix “Tax/” in Bucket “Finance” expire after an age value of 180 days.

- **Transition**: transition objects automatically – objects in Bucket named “2017” move on last day of year from original storage location in Scality RING named “RING-Primary” to Amazon S3 in the US-West (N. California) under the location name “S3-West.”

- **1:1 Replication**: replicate (copy) objects – objects from “ContentHome” with location in Scality RING named “RING-Primary” copied to Amazon S3 Bucket in US-East (Ohio) region in a Bucket named “Content-Delivery-1.”

- **1:Many Replication**: replicate (copy) objects to **multiple locations simultaneously with automatic status alerts and re-tries** - objects from “ContentHome” with location in Scality RING named “RING-Primary” copied to Amazon S3 Bucket in US-East (Ohio) region in a Bucket named “Content-Delivery-1” and copied to Azure Blob Storage Container in North Central US named “Content-Delivery-2” and copied to Google Cloud Storage Bucket in Belgium named “Content-Delivery-3.”

Policies can have granularity applied to them such that when replicating objects, for example, the policy would only replicate objects whose prefix was “Videos/” after objects became more than seven days old.
XI. Metadata Search

Zenko manages data across all of its storage locations, including public clouds, private clouds, and other on-premises storage locations. A core aspect of Zenko is that it stores metadata attributes in addition to maintaining location information (keys) for the object (data) payloads, for purposes of lookup and policy-based workflows. In addition, a global search capability to enable objects to be retrieved based on metadata attribute values, as described here.

Zenko, by default, captures a set of system-defined metadata attributes for objects written through Zenko, and stores them in its Metadata database. System-defined metadata includes common attributes such as the object owner id, create time, object size in bytes, location, keys, and more. This alone allows applications and Zenko Orbit users to retrieve objects based on age, size, cloud location or a combination of these and other search criteria.

In addition, Zenko enables applications and end-users to create extended (application or user-defined) metadata attributes to be associated with each object. The underlying Amazon S3 API provides an extensible metadata capability that allows these attributes to be applied to objects at the time of an Object PUT operation, with additional API (“x-amz-meta-”) headers. Up to 2Kbytes of key/value pairs of metadata storage is provided per object, in adherence to the S3 API specification.

For example, a Media Asset Management (MAM) application may require the storage of extended attributes on video files, such as contentAuthor, contentType, contentLength, contentFirstAired, contentStarring, and other attributes of various data types. Zenko then provides the ability for the application to perform searches based on these key/values and applications to retrieve objects based on attribute values. An example of this would be retrieving all video files created on a certain date, with a specified type and Author. The search criteria can include both system and extended/user-defined attributes in a single search.

Zenko enables metadata searches through an extension of the standard AWS S3 “GET Bucket” API, which makes retrieval of object keys based on such attribute values possible. API searches can utilize the standard GET Bucket API to retrieve a listing of all the object keys in a bucket. By appending a search conditional string as an additional header on the Get
Object API, this will retrieve only the keys of the objects matching the search. The search conditional string syntax is as flexible as SQL based searches in relational database systems (the "where" clause of an SQL search is very similar). This means that single or multi-attribute searches, multiple clauses connected with “AND” and “OR” criteria, and usage all of the normal equality/inequality and less-than/greater-than matching operators. For string based searches, Zenko also supports then concept of “fuzzy” searches (similar to “LIKE” in SQL), and partial condition matches using a wildcard in the string to be matched.

The Zenko Orbit portal also provides a simple user interface for metadata attribute creation, through the Multicloud object browser, as well as interactive Metadata search through a dedicated page in the UI.

By enabling search on both application-defined and system metadata, across the global set of all underlying clouds and RINGs, Zenko provides a very powerful mechanism for enriching the capabilities of a wide range of applications. By unifying this under the well-known S3 API, this makes this new capability a simple extension to applications already using S3.
XII. Cloud-based Management Portal : Zenko Orbit

Zenko Orbit is a cloud-based portal that simplifies the management of Zenko's multi-cloud capabilities for data placement, workflows and global search. Orbit is the easiest way to get started with Zenko to manage multi-cloud storage. Orbit also helps analyze your cloud data usage with detailed statistics, and monitoring of the Zenko instance itself.

Through a point-and-click graphical interface, Orbit provides an easy way for cloud application developers to create Zenko workflows such as replication, lifecycle management and easy data upload/download. Orbit’s simple interface includes the ability to search across the single, global namespace on system and user-defined metadata attributes.
Summary

The advantages of multi-cloud are great, and Scality is delivering on these promises today with the Zenko multi-cloud data controller. Zenko enables a robust solution for efficient Hybrid cloud storage, with powerful policy-based data management for lifecycle and replication, federated search and support for cloud-based applications using the S3 protocol. It opens-up immense efficiency and opportunities for leveraging and sharing data by separating metadata from the underlying data storage in the cloud and ensuring compatibility of existing clouds apps and services through native storage formats for all data stored in the cloud.

Get started today with a free trial of Zenko at www.zenko.io.